



# Optimisation en présence de contraintes en probabilité et processus markoviens contrôlés

Mauricio Gonzalez, Patricia Bouyer, Samson Lasaulce, Nicolas Markey

## ► To cite this version:

Mauricio Gonzalez, Patricia Bouyer, Samson Lasaulce, Nicolas Markey. Optimisation en présence de contraintes en probabilité et processus markoviens contrôlés. GRETSI 2019 - XXVIIème Colloque GRETSI Traitement du Signal & des Images, Aug 2019, Lille, France. pp.1-4. hal-02268161

**HAL Id: hal-02268161**

**<https://hal.science/hal-02268161>**

Submitted on 20 Aug 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Optimisation en présence de contraintes en probabilité et processus markoviens contrôlés \*

Mauricio GONZÁLEZ<sup>1,2</sup>, Patricia BOUYER<sup>1</sup>, Samson LASAULCE<sup>2</sup>, Nicolas MARKEY<sup>3</sup>

<sup>1</sup>CNRS, LSV, ENS Paris-Saclay. 61, avenue du Président Wilson, 94235 Cachan cedex, France.

<sup>2</sup>CNRS, L2S, CentraleSupélec. 3, rue Joliot-Curie, 91192 Gif sur Yvette, France.

<sup>3</sup>CNRS, IRISA, U. Rennes 1. 263, avenue du Général Leclerc, 35042 Rennes cedex, France.

mauricio.gonzalez@lsv.fr, patricia.bouyer@lsv.fr, samson.lasaulce@lss.supelec.fr,  
nicolas.markey@irisa.fr

**Résumé** – Cet article s’intéresse à l’existence et la synthèse des stratégies dans les processus de décision de Markov doublement pondérés, qui satisfont à la fois une contrainte en probabilité sur une condition d’accessibilité pondérée, et une contrainte quantitative sur la valeur espérée d’une variable aléatoire définie à l’aide d’une autre condition d’accessibilité. Cette étude généralise un problème d’ordonnancement de stratégies de consommation de puissance dans les réseaux électriques (tel que la recharge de véhicules électriques). Nous étudions l’ensemble des valeurs d’un paramètre (un seuil) pour lequel le problème a une solution, et montrons comment une caractérisation partielle de cet ensemble peut être obtenue via deux séquences de problèmes d’optimisation. Nous discutons également la complétude et la faisabilité de la méthode développée ainsi obtenue.

**Abstract** – This article focuses on the existence and synthesis of strategies in double-weighted Markov decision processes, which satisfy both a probability constraint over a weighted reachability condition, and a quantitative constraint on the expected value of a random variable defined using another reachability condition. This work generalizes a scheduling problem for energy consumption (typically the problem of charging electric vehicles). We study the set of values of a parameter (a threshold) for which the problem has a solution, and show how a partial characterization of this set can be obtained via two sequences of optimization problems. We also discuss the completeness and feasibility of the resulting approach.

## 1 Introduction

Les méthodes formelles peuvent aider à fournir diverses solutions algorithmiques pour la conception des stratégies. Le problème de recharge des appareils électriques contrôlables est un exemple d’un tel domaine d’application. Ce problème, généralement présenté comme un problème de contrôle (e.g., la recharge des Véhicules Électriques (VEs)), peut en fait être modélisé à l’aide d’un Processus de Décision de Markov (MDP) [1]. Probabilités offrent un moyen de modéliser la partie non-contrôlable de la consommation sur le Réseau de Distribution d’énergie (RD) (consommation en dehors les appareils électriques contrôlables), pour laquelle des grandes bases de données existent afin d’extraire des statistiques précises. Nous pouvons ensuite (i) exprimer une limite supérieure de la charge maximale sur le RD (codée comme une condition d’accessibilité dans notre modèle), (ii) une contrainte imposée à la recharge de tous les appareils contrôlables en tant qu’objectif quantitatif d’accessibilité, et (iii) divers critères d’optimisation sur des coûts (e.g., minimiser le vieillissement du transformateur du RD). Motivés par cet exemple, nous développons une technique permettant de calculer une stratégie optimale dans un MDP, reprenant les résultats de [2, 3].

Tout d’abord, nous présentons la modélisation du problème (généralisation au problème d’énergie [1]) et caractérisons l’existence d’une solution (stratégie) à travers des problèmes d’optimisation sous-jacents. Pour résoudre ces derniers, nous introduisons des problèmes duaux associés [4], et montrons comment construire une solution. Comme les valeurs optimales des problèmes initial et dual sont différentes (écart de dualité), en particulier pour des stratégies pures, nous étendons ces problèmes aux stratégies mixtes et prouvons qu’il n’y a pas d’écart de dualité dans un tel contexte. Nous montrons enfin qu’on peut construire une solution mixte à travers de deux stratégies pures et développons un algorithme pour telle construction.

## 2 Contexte

Le problème (centralisé) de la recharge des appareils électriques contrôlables (e.g., des VEs) consiste en l’ordonnement de charges dans un intervalle de temps  $[0, T]$  en présence de charges non-contrôlables incertaines de manière à minimiser l’impact de la charge totale sur le RD. Une première contrainte sur le transformateur est donnée par sa capacité totale à chaque instant, et une deuxième peut représenter la recharge requise (e.g., recharger les VEs). La charge contrôlable peut donc être aussi considérée comme une première fonction de coût. En utili-

\*Remerciements au projet ERC EQualIS, à la chaire RTE-CentraleSupélec.

sant un modèle standard pour le vieillissement du transformateur (e.g., voir [1]), nous pouvons exprimer ce dernier sous la forme d'une deuxième fonction de coût. Ainsi, un MDP doublement pondéré peut être construit. Dans ce contexte, nous combinons (voir Déf. 2.2) une condition d'accessibilité sûre vers un ensemble cible (en représentant, e.g., la recharge des VEs), une contrainte en probabilité sur la proportion de chemins ayant une valeur supérieure à certain seuil avec la première fonction de coût (e.g., contraint sur la charge totale sur le RD), et une contrainte sur la valeur espérée d'une fonction du deuxième coût (e.g., le vieillissement du transformateur).

**Définition 2.1 (MDP pondéré)** *Un Processus de Décision de Markov pondéré est un 5-uplet  $\mathcal{M} := (\mathcal{S}, s_0, \mathcal{A}, \delta, (c_j)_{j=1}^J)$ , avec  $J \in \mathbb{N}$ , où  $\mathcal{S}$  est un ensemble fini d'états,  $s_0 \in \mathcal{S}$  un état initial,  $\mathcal{A}$  un alphabet fini d'« actions »,  $\delta: \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{D}(\mathcal{S})$ , avec  $\mathcal{D}(\mathcal{S}) := \{\mu: \mathcal{S} \rightarrow [0, 1] \mid \sum_{s \in \mathcal{S}} \mu(s) = 1\}$  l'ensemble de distributions sur  $\mathcal{S}$ , et  $c_j: \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$  pour chaque  $j = 1, \dots, J$ .*

Nous notons par  $\Omega_{s_0}$  (resp.  $\underline{\Omega}_{s_0}$ ) l'ensemble des chemins infinis  $\omega := s_0 s_1 \dots$  (resp. finis) à partir de  $s_0$  dans  $\mathcal{M}$ . Pour  $T \in \mathbb{N}$ ,  $\omega_T$  dénote un chemin de longueur  $T$ . Un  $\mathcal{T} \subseteq \mathcal{S}$  est un **ensemble cible** si nous voulons atteindre un état de  $\mathcal{T}$  à partir de  $s_0$ , en utilisant une **stratégie** (une fonction qui nous permet de choisir les actions). Une stratégie est **pure** si elle est de la forme  $\pi: \underline{\Omega}_{s_0} \rightarrow \mathcal{A}$ , et **mixte** si elle assigne une probabilité sur les stratégies pures, c.-à-d.,  $\sigma: \Pi \rightarrow [0, 1]$  telle que  $\sum_{\pi \in \Pi} \sigma(\pi) = 1$ . Nous considérons par la suite l'ensemble des stratégies pures utilisant une mémoire bornée (sur les chemins). Cet ensemble, noté  $\Pi$ , est fini, une fois fixée la borne de mémoire. Nous disons qu'une stratégie est **sans mémoire** si elle ne dépend que de l'état courant de  $\mathcal{M}$  et non des états visités précédemment. Nous notons l'ensemble des stratégies mixtes par  $\Delta[\Pi]$  et définissons  $K = |\Pi|$  le cardinal de  $\Pi$ . Quand une stratégie est fixée, les chemins résultants dans  $\mathcal{M}$  sont complètement aléatoires, où aucune décision n'est plus prise. Ainsi,  $\Omega_{s_0}^\pi$  (resp.  $\Omega_{s_0}^\sigma$ ) dénote l'ensemble des chemins générés par une stratégie pure  $\pi$  (resp. mixte  $\sigma$ ). En outre, la distribution de probabilité selon  $\pi$  (resp.  $\sigma$ ) sur les chemins infinis, est notée par  $\mathbb{P}_{s_0}^\pi$  (resp.  $\mathbb{P}_{s_0}^\sigma$ ). Pour une fonction  $F$ , sa valeur espérée peut être ainsi calculée selon  $\mathbb{P}_{s_0}^\pi$  (resp.  $\mathbb{P}_{s_0}^\sigma$ ), et elle sera notée par  $\mathbb{E}_{s_0}^\pi[F]$  (resp.  $\mathbb{E}_{s_0}^\sigma[F]$ ).

Nous nous intéressons aux propriétés d'accessibilité quantitative, également appelée Somme Tronquée (ST), pour atteindre un  $\mathcal{T} \subset \mathcal{S}$ . Pour  $\omega \in \Omega_{s_0}$  et  $1 \leq j \leq J$ , nous définissons la somme tronquée par  $ST_j^T(\omega) = \sum_{t=0}^{T_j} c_j(s_t, a_t)$  si  $\omega$  atteint  $\mathcal{T}$ , et  $ST_j^T(\omega) = +\infty$  sinon; dans cette définition,  $T_j$  est le premier indice de temps où  $\omega$  atteint  $\mathcal{T}$ . Motivés par toutes les considérations pratiques, le problème qui nous intéresse peut alors être défini.

## 2.1 Position du problème

Le problème que nous abordons dans cet article provient d'une récente étude d'un problème de recharge de véhicules électriques [1]. Puisque les stratégies pures sont vues comme un cas particulier des stratégies mixtes, nous définissons le problème pour ces dernières.

**Définition 2.2 (Le Problème)** *Étant donné un MDP pondéré  $\mathcal{M}$ , soit  $\mathcal{T} \subset \mathcal{S}$  un ensemble cible et  $\nu_1, \nu_2 \in \mathbb{R}$  deux seuils. Pour chaque  $\varepsilon \in [0, 1]$ , le problème  $\mathbf{Pb}(\varepsilon)$  est défini comme : il existe  $\sigma_\varepsilon$  telle que pour chaque  $\omega \in \Omega_{s_0}^{\sigma_\varepsilon}$ ,  $\omega$  atteint  $\mathcal{T}$  et,*

$$(i) \mathbb{P}_{s_0}^{\sigma_\varepsilon}[ST_1^T \geq \nu_1] \geq 1 - \varepsilon \quad \text{et} \quad (ii) \mathbb{E}_{s_0}^{\sigma_\varepsilon}[ST_2^T] \leq \nu_2$$

Notre objectif est de calculer les valeurs de  $\varepsilon$  pour lequel  $\mathbf{Pb}(\varepsilon)$  a une solution. Nous supposons qu'il existe une stratégie satisfaisant la contrainte (ii). Sinon,  $\mathbf{Pb}(\varepsilon)$  n'a aucune solution quel que soit  $\varepsilon$ . Ce dernier problème est réduit au problème du « plus court chemin », où l'existence d'une stratégie  $\delta$ -optimale, pure et sans mémoire  $\pi_\delta^*$ , telle que  $\mathbb{E}_{s_0}^{\pi_\delta^*}[ST_2^T] \leq \nu_2 + \delta$  avec  $\delta > 0$ , peut être décidée en temps polynomiale [5].

Dans ce qui suit, nous appellerons **cartographie** de  $\mathbf{Pb}(\varepsilon)$ , la fonction qui associe à chaque  $\varepsilon$ , soit VRAI si  $\mathbf{Pb}(\varepsilon)$  a une solution, soit FAUX sinon. Nous décrivons un algorithme pour cartographier approximativement  $\mathbf{Pb}(\varepsilon)$  sur l'intervalle  $[0, 1]$ .

## 3 Cartographie approximée

Nous introduisons deux problèmes d'optimisation liés à  $\mathbf{Pb}(\varepsilon)$ , à partir desquels nous obtenons des informations sur les valeurs de  $\varepsilon$  pour lesquelles  $\mathbf{Pb}(\varepsilon)$  a une solution. Notre approche caractérise partiellement l'intervalle  $[0, 1]$ , cependant sous certaines conditions dans la structure de  $\mathcal{M}$ , la cartographie est presque complète [2].

Pour  $T \in \mathbb{N}$ , nous définissons deux événements sur les chemins  $\omega_T: \mathcal{E}_T := \{\omega_T \in \underline{\Omega}_{s_0} \mid \omega_T \text{ atteint } \mathcal{T}\}$  et,  $\bar{\mathcal{E}}_{\alpha, T} := \{\omega_T \in \underline{\Omega}_{s_0} \mid \omega_T \text{ n'atteint pas } \mathcal{T}\}$  quand  $\alpha = 1$  et  $\bar{\mathcal{E}}_{\alpha, T} := \emptyset$  quand  $\alpha = 0$ .

### 3.1 Problèmes d'optimisation

Ici, nous considérons une **fonction objective** pour  $T \in \mathbb{N}$  fixé. Pour une stratégie pure  $\pi \in \Pi$ , elle est définie par :

$$\mathbb{J}_{\alpha, T}(\pi) := \mathbb{P}_{s_0}^\pi[(\mathcal{E}_T \cap ST_1^T \leq \nu_1) \cup \bar{\mathcal{E}}_{\alpha, T}] \quad (1)$$

Ainsi, les deux problèmes d'optimisation peuvent être regroupés sous le paramètre  $\alpha \in \{0, 1\}$  comme :

$$\min_{\pi} \{ \mathbb{J}_{\alpha, T}(\pi) \mid \mathbb{E}_{s_0}^\pi[ST_2^T] \leq \nu_2 \} \quad (2)$$

Puisque les stratégies mixtes sont vues comme des combinaisons convexes des stratégies pures, il est facile de voir que (2) peut être écrit pour des stratégies mixtes comme :

$$\min_{\sigma} \left\{ \sum_{k=1}^K \sigma(\pi^k) \mathbb{J}_{\alpha, T}(\pi^k) \mid \sum_{k=1}^K \sigma(\pi^k) \mathbb{E}_{s_0}^{\pi^k}[ST_2^T] \leq \nu_2 \right\} \quad (3)$$

Nous écrivons la **valeur optimale** de (2) (resp. de (3)) comme  $\mathbb{J}_{\alpha, T}^{P*}$  (resp.  $\mathbb{J}_{\alpha, T}^{M*}$ ). Notez que  $\mathbb{J}_{\alpha, T}^{M*} \leq \mathbb{J}_{\alpha, T}^{P*}$ , car des stratégies mixtes nous permettent d'obtenir une valeur optimale qui pourrait être meilleure que sous des stratégies pures. Pour les deux contextes des stratégies, nous notons simplement  $\mathbb{J}_{\alpha, T}^*$ . Il n'est pas difficile de voir que pour chaque  $T \in \mathbb{N}$ ,  $\mathbb{J}_{\alpha, T}^* \leq \mathbb{J}_{\alpha, T+1}^*$ ; et que la séquence  $(\mathbb{J}_{\alpha, T}^*)_{T \in \mathbb{N}}$  est croissante pour  $\alpha = 0$  et décroissante pour  $\alpha = 1$ . Puisque cette dernière est bornée, elle est donc convergente [2]. Nous écrivons sa limite par  $\mathbb{J}_{\alpha, \infty}^* := \lim_{T \rightarrow +\infty} \mathbb{J}_{\alpha, T}^*$ .

### 3.2 Approche presque complète

Les deux lemmes suivants caractérisent la cartographie approximative du problème  $\text{Pb}(\varepsilon)$ .

**Lemme 3.1 ([2])**  $\forall \varepsilon < \mathbb{J}_{1,\infty}^*$ ,  $\text{Pb}(\varepsilon)$  n'a pas de solution.

**Lemme 3.2 ([2])** Pour chaque  $T \in \mathbb{N}$ ,  $\text{Pb}(\varepsilon)$  a une solution  $\forall \varepsilon > \mathbb{J}_{1,T}^* \geq \mathbb{J}_{1,\infty}^*$ , et n'a pas de solution  $\forall \varepsilon < \mathbb{J}_{0,T}^* \leq \mathbb{J}_{0,\infty}^*$ .

La Figure 1 résume l'analyse des lemmes précédents. Théoriquement (Lemme 3.1), elle est presque-complète, puisque le seul statut de  $\varepsilon = \mathbb{J}_{1,\infty}^*$  reste incertain. Cependant, il reste à discuter une chose pour faire un calcul effectif (Lemme 3.2) : les limites  $\mathbb{J}_{0,\infty}^*$  et  $\mathbb{J}_{1,\infty}^*$  sont à priori inconnues par calcul, d'où la cartographie n'est pas efficace. L'idée est alors d'utiliser la séquence  $(\mathbb{J}_{\alpha,T}^*)_{T \in \mathbb{N}}$  pour chaque  $\alpha$ , pour se rapprocher aux limites. Cependant, si les deux limites coïncident, nous aurions une approche presque complète et donc un algorithme couvrant presque l'intervalle  $[0, 1]$  avec soit la ligne rouge (où il n'y a pas de solution), soit la ligne verte (où il y a une solution). Il existe des situations (relatives aux cycles dans  $\mathcal{M}$ ) dans lesquelles on peut montrer que  $\mathbb{J}_{0,\infty}^* = \mathbb{J}_{1,\infty}^*$ , ce qui permet de réduire la partie inconnue de la cartographie à un seul singleton.

**Lemme 3.3 ([2])** Si tous les cycles ont des coûts positifs<sup>1</sup> sous  $c_j$ , pour au moins un  $j \in \{1, 2\}$ ; alors  $\mathbb{J}_{0,\infty}^* = \mathbb{J}_{1,\infty}^*$ .

Il nous reste seulement à savoir comment résoudre les problèmes d'optimisation, pour que l'algorithme soit complète. C'est l'un des objectifs de la section suivante.

## 4 Solution aux problèmes d'optimisation

Pour  $T \in \mathbb{N}$  fixé, nous déplaçons  $\mathcal{M}$  jusqu'à la profondeur  $T$  comme un arbre, en gardant une copie de  $\mathcal{M}$  sous chaque feuille. À partir de chacune, nous utiliserons toujours la stratégie du plus court chemin  $\pi_\delta^*$ , de sorte que les problèmes d'optimisation sont réduits à trouver une solution  $\forall t < T$ . Nous notons  $\bar{\pi}_\delta := \pi \pi_\delta^*$  pour une stratégie qui joue depuis  $t = 0$  jusqu'à  $t = T - 1$  comme une stratégie pure  $\pi$  sur l'arbre, et comme  $\pi_\delta^*$  pour chaque  $t \geq T$ . Les solutions des problèmes d'optimisation sont obtenues en résolvant des problèmes duaux sous-jacents.

### 4.1 Approche Lagrangienne

Nous définissons la **fonction Lagrangienne** pour des stratégies pures  $\bar{\pi}_\delta$  et pour un paramètre  $\lambda \in \mathbb{R}^+$  (appelé **variable duale**), comme :

$$\mathbb{L}_{\alpha,T}(\bar{\pi}_\delta, \lambda) := \mathbb{J}_{\alpha,T}(\bar{\pi}_\delta) + \lambda (\mathbb{E}_{s_0}^{\bar{\pi}_\delta} [\text{ST}_2^T] - \nu_2) \quad (4)$$

Comme dans (3), il n'est pas difficile de voir que la version de (4) pour des stratégies mixtes, peut être exprimée par :

$$\mathbb{L}_{\alpha,T}(\sigma, \lambda) = \sum_{k=1}^K \sigma(\bar{\pi}_\delta^k) \mathbb{L}_{\alpha,T}(\bar{\pi}_\delta^k, \lambda) \quad (5)$$

1. Lorsque nous supposons que les cycles ont des coûts positifs, nous le pensons pour chaque cycle dans  $\mathcal{M}$ .

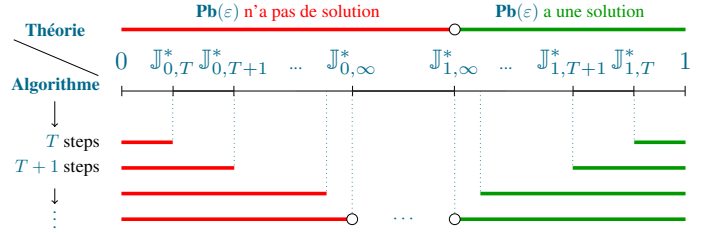


FIGURE 1 – Cartographie partielle du problème.

En prenant le minimum sur des stratégies pures (resp. mixtes) dans (4) (resp. (5)), nous dénotons resp. la **fonction Lagrangienne dual** pour des stratégies pures et mixtes, comme :

$$\mathbb{L}_{\alpha,T}^p(\lambda) := \min_{\bar{\pi}_\delta} \mathbb{L}_{\alpha,T}(\bar{\pi}_\delta, \lambda), \quad \mathbb{L}_{\alpha,T}^m(\lambda) := \min_{\sigma} \mathbb{L}_{\alpha,T}(\sigma, \lambda) \quad (6)$$

Il se trouve que les fonctions (4) et (5) sont toujours concaves, et ainsi, les fonctions (6) aussi [4]. Nous définissons resp. le **problème dual** pour des stratégies pures et mixtes, par :

$$\mathbb{L}_{\alpha,T}^{p*} := \sup_{\lambda \geq 0} \mathbb{L}_{\alpha,T}^p(\lambda), \quad \mathbb{L}_{\alpha,T}^{m*} := \sup_{\lambda \geq 0} \mathbb{L}_{\alpha,T}^m(\lambda) \quad (7)$$

Dans ce contexte, le « théorème de dualité faible » peut s'appliquer [4], montrant que  $\mathbb{J}_{\alpha,T}^{p*} - \mathbb{L}_{\alpha,T}^{p*} \geq 0$  et  $\mathbb{J}_{\alpha,T}^{m*} - \mathbb{L}_{\alpha,T}^{m*} \geq 0$ . Quand l'inégalité est stricte, nous disons qu'il n'existe qu'une **dualité faible**, et si  $\mathbb{J}_{\alpha,T}^* = \mathbb{L}_{\alpha,T}^*$ , nous disons qu'il y a une **dualité forte**. On peut montrer qu'il y a une dualité faible sous des stratégies pures et une dualité forte sous des stratégies mixtes. Ainsi, nous pouvons résoudre (7) pour des stratégies pures et donc, la variable dual optimale  $\lambda^*$  sera aussi une solution au problème dual pour des stratégies mixtes. Puisqu'il y a une dualité forte sous  $\Delta[\Pi]$ , trouver une solution à (3) est équivalent à trouver une  $\sigma_\alpha^*$  optimale à (6) avec  $\lambda^*$  fixée. Ce dernier est un problème sans contrainte et puisque pour chaque  $t \geq T$  nous utilisons la stratégie  $\pi_\delta^*$ , le problème est réduit à une minimisation dans le MDP déplié jusqu'à  $T$ .

**Théorème 4.1** Soit  $\sigma_\alpha^*$  et  $\lambda^*$  faisables resp. pour le problème d'optimisation (3) et le problème dual des stratégies mixtes (7). Supposons que  $\sigma_\alpha^* \in \arg \min_{\sigma} \mathbb{L}_{\alpha,T}(\sigma, \lambda^*)$  et que

$$\lambda^* (\mathbb{E}_{s_0}^{\sigma_\alpha^*} [\text{ST}_2^T] - \nu_2) = 0 \quad (8)$$

Alors,  $\sigma_\alpha^*$  et  $\lambda^*$  sont optimales, et  $\mathbb{L}_{\alpha,T}^{m*} = \mathbb{L}_{\alpha,T}^m(\lambda^*) = \mathbb{L}_{\alpha,T}^{p*} = \mathbb{L}_{\alpha,T}^p(\lambda^*) = \mathbb{J}_{\alpha,T}^{m*} = \mathbb{J}_{\alpha,T}^m(\sigma_\alpha^*) \leq \mathbb{J}_{\alpha,T}^{p*}$ .

La preuve est omise. Notez que si  $\lambda^* = 0$ , alors la contrainte n'est plus active, c.-à.-d., une stratégie pure résolvant le problème sans contrainte est une solution au problème (2) et des stratégies mixtes ne sont pas nécessaires comme indiqué dans la Figure 2 à droite, car  $\mathbb{J}_{\alpha,T}^* = \mathbb{L}_{\alpha,T}^p(0)$ . La proposition suivante sera utile pour construire  $\sigma_\alpha^*$  si c'est le cas.

**Proposition 4.2 ([3])** Soit  $\bar{\pi}_{\delta,\alpha,\lambda} \in \arg \min_{\bar{\pi}_\delta} \mathbb{L}_{\alpha,T}(\bar{\pi}_\delta, \lambda)$  pour chaque  $\lambda \geq 0$ . Alors,  $(\mathbb{E}_{s_0}^{\bar{\pi}_{\delta,\alpha,\lambda}} [\text{ST}_2^T])_{\lambda \geq 0}$  est décroissante.

Finalement, nous montrons le genre de stratégie qui résout le problème (3).

**Théorème 4.3** Une solution au problème (3) pour  $\alpha$  fixé, est une stratégie mixte construite d'au plus deux stratégies pures.

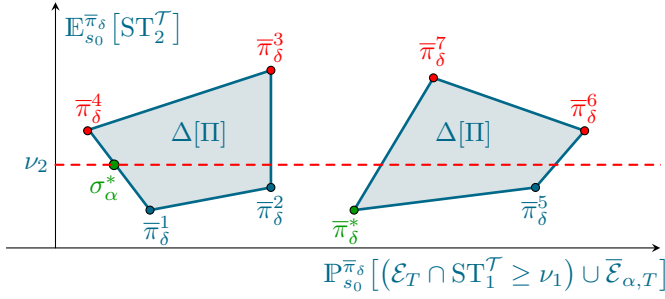


FIGURE 2 – Sur la gauche,  $\sigma_\alpha^*$  est optimale, combinant les stratégies pures  $\pi_\delta^1$  et  $\pi_\delta^4$ . Sur la droite,  $\pi_\delta^*$  est optimale (solution du problème sans contrainte).

La preuve est omise. La Figure 2 montre une représentation graphique du Théorème 4.3. Nous examinons l’Algorithme 1 pour expliquer ce théorème. Tout d’abord, nous résolvons (2) sans contrainte (cas  $\lambda^* = 0$ ) et calculons la valeur de l’espérance dans la contrainte. Si elle est inférieure à  $\nu_2$ , alors une stratégie pure est une solution (lignes 2-3). En supposant que  $\lambda^* > 0$ , on s’approche vers  $\lambda^*$  à travers la méthode de dichotomie, en générant deux suites  $(\lambda_n)_{n \in \mathbb{N}}$  et  $(\bar{\lambda}_n)_{n \in \mathbb{N}}$ , telles que  $\lambda_n \nearrow \lambda^*$  et  $\bar{\lambda}_n \searrow \lambda^*$ . Nous initialisons les suites comme  $\lambda_0 := 0$  et  $\bar{\lambda}_0 := \lambda^+ > 0$ . Si  $\lambda^+$  n’est pas assez grande, nous cherchons une autre  $\lambda^+$  jusqu’à avoir une stratégie pure qui satisfasse la contrainte. (lignes 4-8). Ainsi, on aura  $\pi_{\lambda_0}$  qui satisfait la contrainte et  $\pi_{\bar{\lambda}_0}$  qui ne la satisfait pas. Ensuite, on passe à la construction d’une stratégie mixte  $\sigma_\alpha^*$  (lignes 10-23). Pour chaque  $\lambda_n$  et  $\bar{\lambda}_n$  fixés, on calcule  $\pi_{\lambda_n}$  et  $\pi_{\bar{\lambda}_n}$  à travers de la récursion classique de Bellman [6], en minimisant (4). Cette approche générera deux séquences monotones de valeurs espérées de la contrainte (Proposition 4.2). Puisque par la méthode de dichotomie,  $\lambda_n = (\bar{\lambda}_{n-1} + \lambda_{n-1})/2$  est le point médian de l’intervalle  $[\lambda_{n-1}, \bar{\lambda}_{n-1}]$  à l’itération  $n$ , on peut montrer que  $|\lambda_n - \lambda^*| \leq (\bar{\lambda}_0 - \lambda_0)/2^{n-1}$ , et alors que  $|\mathbb{J}_{\alpha, T}^m(\sigma_\alpha^*) - \mathbb{J}_{\alpha, T}^m| \leq M_n(\bar{\lambda}_0 - \lambda_0)/2^{n-1}$ , où  $M_n$  est en fonction des valeurs espérées avec  $\pi_{\lambda_n}$  et  $\pi_{\bar{\lambda}_n}$  à l’itération  $n$  (ligne 21). Nous itérons jusqu’à ce que la dernière inégalité soit inférieure à un erreur  $\epsilon$ . Ainsi, nous aurons  $\pi_{\lambda_n}$  et  $\pi_{\bar{\lambda}_n}$  pour effectuer une combinaison convexe entre elles, en résolvant (8) avec  $\lambda^* > 0$ . Dans ce cas,  $\mathbb{E}_{s_0}^{\pi_{\bar{\lambda}_n}}[ST_2^T] < \nu_2 < \mathbb{E}_{s_0}^{\pi_{\lambda_n}}[ST_2^T]$  et, puisque l’espérance sous des stratégies mixtes est continue, il existera une  $\sigma_\alpha^*$  telle que  $\sigma_\alpha^*(\pi_{\lambda_n}) + \sigma_\alpha^*(\pi_{\bar{\lambda}_n}) = 1$ , et  $\sigma_\alpha^*(\pi_{\lambda_n})\mathbb{E}_{s_0}^{\pi_{\lambda_n}}[ST_2^T] + \sigma_\alpha^*(\pi_{\bar{\lambda}_n})\mathbb{E}_{s_0}^{\pi_{\bar{\lambda}_n}}[ST_2^T] = \nu_2$  (ligne 23).

## 5 Conclusion

Bien que le problème théorique abordé dans cet article ait été inspiré du problème d’ordonnancement de la consommation de puissance dans les réseaux électriques intelligents, sa formulation et ses solutions concernent une classe bien plus large de problèmes pratiques apparaissant en traitement du signal. Nous pensons en particulier au problème de contrôle de puissance dans les communications sans fil avec des contraintes de qualité de service (probabilité de coupure typiquement).

**Entrée :**  $\alpha \in \{0, 1\}$ ,  $T \in \mathbb{N}$ ,  $\lambda^+ > 0$ ,  $\epsilon > 0$ ,  $\pi_\delta^*$ .  
**Sortie :** Solution au Problème (3).

```

1  $(\lambda_0, \bar{\lambda}_0) \leftarrow (0, \lambda^+)$ 
2  $\pi_{\lambda_0} \leftarrow \pi_{\delta, \alpha}^* \in \arg \min_{\pi_\delta} \mathbb{J}_{\alpha, T}(\pi_\delta)$ 
3 si  $\mathbb{E}_{s_0}^{\pi_{\lambda_0}}[ST_2^T] \leq \nu_2$  alors, retourner  $\pi_{\lambda_0}$ ;
4  $\pi_{\bar{\lambda}_0} \leftarrow \pi_{\delta, \alpha, \bar{\lambda}_0}^* \in \arg \min_{\pi_\delta} \mathbb{L}_{\alpha, T}(\pi_\delta, \bar{\lambda}_0)$ 
5 tant que  $\mathbb{E}_{s_0}^{\pi_{\bar{\lambda}_0}}[ST_2^T] \geq \nu_2$  faire
6   si  $\mathbb{E}_{s_0}^{\pi_{\bar{\lambda}_0}}[ST_2^T] = \nu_2$  alors, retourner  $\pi_{\bar{\lambda}_0}$ ;
7    $\bar{\lambda}_0 \leftarrow \lambda^+ > \bar{\lambda}_0$ 
8    $\pi_{\bar{\lambda}_0} \leftarrow \pi_{\delta, \alpha, \bar{\lambda}_0}^* \in \arg \min_{\pi_\delta} \mathbb{L}_{\alpha, T}(\pi_\delta, \bar{\lambda}_0)$ 
9  $(n, M_0) \leftarrow (0, \epsilon/(\bar{\lambda}_0 - \lambda_0))$ 
10 tant que  $\log_2(2M_n(\bar{\lambda}_0 - \lambda_0)/\epsilon) > n$  faire
11    $n \leftarrow n + 1$ 
12    $\lambda_n \leftarrow (\bar{\lambda}_{n-1} + \lambda_{n-1})/2$ 
13    $\pi_{\lambda_n} \leftarrow \pi_{\delta, \alpha, \lambda_n}^* \in \arg \min_{\pi_\delta} \mathbb{L}_{\alpha, T}(\pi_\delta, \lambda_n)$ 
14   cas où  $\mathbb{E}_{s_0}^{\pi_{\lambda_n}}[ST_2^T] = \nu_2$  faire, retourner  $\pi_{\lambda_n}$ ;
15   cas où  $\mathbb{E}_{s_0}^{\pi_{\lambda_n}}[ST_2^T] > \nu_2$  faire
16      $(\lambda_n, \bar{\lambda}_n) \leftarrow (\lambda_n, \bar{\lambda}_{n-1})$ 
17      $(\pi_{\lambda_n}, \pi_{\bar{\lambda}_n}) \leftarrow (\pi_{\lambda_n}, \pi_{\bar{\lambda}_{n-1}})$ 
18   cas où  $\mathbb{E}_{s_0}^{\pi_{\lambda_n}}[ST_2^T] < \nu_2$  faire
19      $(\lambda_n, \bar{\lambda}_n) \leftarrow (\lambda_{n-1}, \lambda_n)$ 
20      $(\pi_{\lambda_n}, \pi_{\bar{\lambda}_n}) \leftarrow (\pi_{\lambda_{n-1}}, \pi_{\lambda_n})$ 
21    $M_n \leftarrow \frac{(\mathbb{E}_{s_0}^{\pi_{\lambda_n}}[ST_2^T] - \nu_2)(\nu_2 - \mathbb{E}_{s_0}^{\pi_{\bar{\lambda}_n}}[ST_2^T])}{\mathbb{E}_{s_0}^{\pi_{\lambda_n}}[ST_2^T] - \mathbb{E}_{s_0}^{\pi_{\bar{\lambda}_n}}[ST_2^T]}$ 
22    $(\sigma_\alpha^*(\pi_{\lambda_n}), \sigma_\alpha^*(\pi_{\bar{\lambda}_n})) \leftarrow \left( \frac{M_n}{\mathbb{E}_{s_0}^{\pi_{\lambda_n}}[ST_2^T] - \nu_2}, \frac{M_n}{\nu_2 - \mathbb{E}_{s_0}^{\pi_{\bar{\lambda}_n}}[ST_2^T]} \right)$ 
23 retourner  $\sigma_\alpha^*$ 

```

**Algorithme 1 :** On initialise avec le problème d’optimisation  $\alpha$  et la profondeur  $T$  du dépliage de  $\mathcal{M}$ , une variable duale initiale  $\lambda^+$ , un erreur  $\epsilon$  fixé, et une stratégie du plus court chemin  $\pi_\delta^*$ .

## Références

- [1] M. González, O. Beaude, P. Bouyer, S. Lasaulce, et N. Markey. *Stratégies d’ordonnancement de consommation d’énergie en présence d’information imparfaite de prévision*. Grets. 2017.
- [2] P. Bouyer, M. González, N. Markey, et M. Randour. *Multi-weighted Markov decision processes with reachability objectives*. arXiv preprint arXiv :1809.03107, 2018.
- [3] F.J. Beutler, et K.W. Ross. *Optimal policies for controlled Markov chains with a constraint*. Journal of mathematical analysis and applications, 112.1 (1985) : 236-252.
- [4] D. Bertsekas. *Convex optimization theory*. Belmont : Athena Scientific, 2009.
- [5] C. Baier, N. Bertrand, C. Dubsiaff, D. Gburek, et O. Sankur. *Stochastic shortest paths and weight-bounded properties in Markov decision processes*. In : LICS’18, IEEE.
- [6] R.E. Bellman, and E.D. Stuart. *Applied dynamic programming*. Vol. 2050. Princeton university press, 2015.